THE EFFECT OF CONSTRUCTING MULTIPLE-CHOICE DISTRACTOR ITEMS AROUND A SINGLE TARGET ALTERNATIVE

Michael S. Wogalter and D. Bradley Marwitz

Department of Psychology, University of Richmond
Richmond, Virginia 23173

ABSTRACT

The present research sought to determine whether the construction of multiple-choice alternatives based around a critical target answer would facilitate the selection of the target answer. Subjects were given a multiple-choice test consisting of 60 questions, each having four alternatives. Twenty of the 60 questions were the critical questions and were constructed to have no correct answer (i.e., asked nonsense) but appeared legitimate. One of the alternatives for the critical questions was the critical alternative, around which the other three distractor alternatives were derived. This was accomplished by systematically substituting each of the critical alternative's three components with another plausible component. This procedure produced a set of alternatives where the critical alternative was more similar to the other alternatives than they were to each other (i.e., it was the most prototypic). The results of two experiments using ranking and proportion scores showed a response bias effect: subjects selected the critical alternatives more often than would be expected by chance. Further analyses revealed that in lower ability subjects the effect disappeared when the critical alternatives were embedded in sets of distractors which had randomly ordered components. High ability subjects selected the critical alternatives more often than chance regardless of the distractors' component arrangement. The results suggest that test-makers should avoid constructing distractor alternatives around a correct alternative because the information provided in the set of alternatives may influence test-takers to select the target answer without any knowledge of the information being assessed.

INTRODUCTION

Prior research has addressed ways to develop better multiple-choice tests and to avoid the pitfalls of test-wiseness on the part of the test-takers (e.g., Metfessel & Sax, 1958; Strang, 1977). Wilcox (1981), for example, suggests that the distractor alternatives should be constructed so as to decrease the probability of the subject guessing the correct alternative. One way to accomplish this is to construct the distractors so that they have some degree of plausibility (Wood, 1960). A technique to ensure plausibility of the distractors is to construct the alternatives so that they are similar to the correct answer. This method of constructing distractors is analogous to procedures used by police departments to construct identification lineups. Lineup members (the distractors) are selected for inclusion in the identification test because of their appearance is relatively similar to the suspect. The problem with this method of construction is that the suspect's appearance becomes *distinctive* in the sense that it becomes more similar to the other lineup members than they are to each other. It is the most *prototypical* face in the lineup. Thus a "witness" may be able to select the police suspect out of a lineup without having seen the suspect before. Wogalter and Jensen (1986) and Laughery, Jensen, and Wogalter (in press) tested this notion by comparing actual selection performance to selection performance expected by chance alone and confirmed the existence of this bias effect for facial stimuli. Moreover, Wogalter and Jensen (1986) have shown the effect for other kinds of pictorial materials as well.

It is unclear from these studies whether the effect is due to the visual-spatial nature of the stimuli employed in these experiments. For example, it is possible that the bias effect occurs only in situations where subjects can compare the visual-spatial appearance of the alternatives which in turn might lead to the production of a prototypic image that is used to match the critical alternative. The present study sought to determine whether the bias effect also holds for non-spatial stimuli, and more specifically, for verbal multiple-choice items. Is there is a tendency or bias for subjects to select a particular multiple-choice alternative when that alternative is more similar to the other alternatives than they are to each other? In other words, can the way the set of distractor alternatives is constructed influence subjects to select a target answer in cases where they have no knowledge of the subject matter being addressed in the stem question?

EXPERIMENT 1

Method

Subjects. Fifty-one University of Richmond undergraduates participated in this study for research credit in an introductory psychology course.

Materials and Stimuli. The testing apparatus was a multiple-choice test consisting of 60 items, each having four alternatives. Of the 60 questions, 40 were legitimate, difficult questions taken from CLEP and GRE preparation manuals. Twenty of the 60 questions were the critical questions and were constructed to have no correct answer while still appearing valid. The question stems for the critical items were derived from real questions from the

preparation manuals but were changed in a some way to ensure what they asked was nonsense (e.g., The flag of the country Candesar contains which 3 colors?). Each critical target alternative was then generated. Each alternative always contained three components (e.g., Yellow, Blue, and White). Each distractor alternative was derived by substituting a different component for each component of the critical target alternative (e.g., Yellow, Blue, and *Black*; *Green*, Blue, and White; Yellow, *Red*, and White). This procedure developed a set of alternatives where the critical alternative was more similar to the other alternatives than they were to each other. That is, the critical alternative was the most prototypic of the alternatives.

Half of the critical questions were arranged in an ordered manner; that is, each component was systematically substituted to maintain the same spatial arrangement as given in the critical alternative, the *ordered component version*. An example set of alternatives appearing as an ordered version is shown below.

a. YELLOW, RED, AND WHITE
b. GREEN, BLUE, AND WHITE
c. YELLOW, BLUE, AND WHITE *
d. YELLOW, BLUE, AND BLACK

Another version of the critical questions contained the same verbal information except the ordering of the components in the distractor alternatives was randomized (scrambled), the *random component version*. An example set of alternatives appearing as an random version is shown below.

a. RED, YELLOW, AND WHITE
b. WHITE, BLUE, AND GREEN
c. YELLOW, BLUE, AND WHITE *
d. BLUE, BLACK, AND YELLOW

The critical questions were randomly interspersed within the set of legitimate questions. Two forms of the test were constructed. One presented the questions in the reverse order of the other test. Both tests contained 10 ordered and 10 random critical questions, but the critical question alternatives that were ordered on one test were random on the other, and vice versa. Thus, the appearance of the ordered or random component versions of each critical question was balanced across test sets.

Procedure. The participants in the experiment were told that the test they were about to take was a difficult general knowledge college-level achievement test. They were told to study each alternative carefully and that they should place the number 1 next to the alternative that seemed most correct to them. In addition they were told to also assign the numbers 2, 3, and 4 to the other alternatives according to their likelihood of being correct with 2 being the next best answer through 4 being the least likely to be correct. Subjects were encouraged to guess if necessary.

Results and Discussion

The data were examined in regard to how often the critical answers were selected relative to what would be expected by random/chance selection. If subjects were merely ranking the question alternatives at random, the mean value for any given alternative should be 2.5 (i.e., subjects assigned the values of 1, 2, 3, and 4 to the set of alternatives for each question, and in the long-run, random assignment of these values to any given alternative would produce an expected mean value of 2.5). Since subjects gave ranking scores, a lower score for the critical alternatives indicates that it was selected more often as being the "correct" answer. In other words, *lower* scores represent *higher* or *greater* selection of the critical alternative.

Selection of the critical alternatives as a function of their appearance in an ordered vs. random set of distractors was examined in three ways: 1) averaging across the critical alternatives to produce mean values for subjects and using these means in the analyses as the random variable, 2) averaging across subjects to produce mean values for the 20 critical alternatives and using these means in the analyses as the random variable, and 3) specific examination of each of the 20 critical questions using the subjects' raw scores in the analyses as the random variable. In all of these analyses the critical alternatives for the ordered and random component versions were compared to the value expected by chance selection. In addition, the difference between the ordered and random component versions was examined.

The first set of analyses used subjects as the random variable which averaged across the critical answers separately for the ordered versions and the random versions. Subjects selected the critical alternative in the ordered version ($M = 2.23$) significantly higher than would be expected by chance selection (2.5), $t\ (50) = 4.38$, $p < .0001$. Similarly, subjects selected the random version ($M = 2.27$) higher than chance, $t\ (50) = 4.06$, $p < .0001$. There was no significant difference between the selection the ordered and random version, $t\ (50) = .68$, $p > .05$.

In the second set of analyses, the subject data was averaged to produce means for the ordered and random critical alternatives for each of the 20 critical questions. These scores were then entered into the analyses as the random variable. Selection of the ordered critical alternatives was significantly better than expected by chance, $t\ (19) = 4.06$, $p < .001$. Similarly, the random versions were selected significantly better than chance, $t\ 19) = 3.76$, $p < .001$. No significant difference between the rankings of the ordered and the random version of the items was found, $t\ (19) = 0.70$, $p > .05$.

In the third set of analyses, each of the 20 critical alternatives was examined individually using the subjects' raw scores as the random variable. Of the 20 critical items, 16 of the ordered critical alternatives were in the direction of being selected more often than expected by chance, and 8 of these were significant ($p's < .05$). Similarly, 16 of the random versions were in the direction of being selected more often than expected by chance, and 6 of these were significant ($p < .05$). Only one out of the 20 would have been expected to differ if chance selection alone was operating.

Thus, the results show that, in spite of the fact that the critical questions addressed nonsense, the prototypic/critical alternative (which is more similar to the other items than they are to each other) is chosen more often than would be expected by chance selection.

Further analyses examined whether higher ability subjects (or better test-takers) perform differently on their selection of the ordered and random critical answers than lower ability subjects. Subjects were divided according to their performance on the legitimate (noncritical) questions using a median split to form two groups of subjects. The two groups, termed *high ability* and *low ability* groups, were compared in regard to their selection of the ordered and random component versions of the critical alternatives. Using subject means for the random and ordered alternatives (averaged across questions), a mixed model ANOVA was employed. No reliable main effects or interaction was noted (p 's > .05). However, positive results were found when selection performance for the four conditions were compared to chance performance. Subjects in the high ability group showed significantly higher than chance selection of the critical alternatives in the ordered versions (M = 2.21) and in the random versions (M = 2.21), t (25) = 3.49, p < .002, and t (25) = 3.93, p < .001, respectively. For the low ability group only the critical alternatives in the ordered versions (M = 2.25) produced selection scores that were significantly different from chance, t (24) = 2.67, p < .02. Selection of the critical alternatives in the random version for the low ability group (M = 2.34) was not significantly different from chance, t (24) = 1.91, p > .05. No difference was found between the ordered and random versions for either the high or low knowledge groups, t (25) = 0.07, p > .05, and t (24) = 0.80, p > .05, respectively. These results suggest that high ability subjects (more adept test-takers) are able to use the information provided in the response alternatives regardless of how the component information is arranged. This use of test material information is suggestive of a form of test-wiseness. Low ability subjects (less adept test-takers) are able to use the information provided by the response alternatives only when the component information is orderly arranged, and not when the component information is randomly arranged (spatially scrambled). Low ability subjects lack some form of test-wiseness that apparently high ability subjects seem to possess.

EXPERIMENT 2

In the first experiment subjects assigned ranking scores to the question alternatives. It is not clear at this point whether the same pattern of results would be found in a task that more closely approximates an actual multiple-choice test. Therefore, in Experiment 2 an attempt was made to gather further support for the response bias effect using a more conventional response mode. Subjects were told to select only a single alternative that best answers each question, rather than assigning rank scores to all alternatives.

Method

Subjects. Fifty-six undergraduate students from the University of Richmond and Virginia Commonwealth University participated in the experiment for extra credit in their psychology class.

Materials and Stimuli. The testing materials were the same as those used in Experiment 1.

Procedure. The participants were told that the test they were about to take was a difficult general knowledge college-level achievement test. They were told to examine each question and its alternatives carefully and to choose the alternative that they felt was correct. The importance of answering every question was stressed, and subjects were encouraged to guess if necessary.

Results and Discussion

As in Experiment 1 the data were examined with regard to how often the critical answers were selected relative to what would be expected by random/chance selection. In this experiment, items were given a score of 1 if subjects selected the prototypic/critical alternative and a 0 if they selected one of the other three alternatives. If subjects were merely selecting at random, then the expected value for the critical alternatives should be one out of four, or 0.25. Since subjects chose only a single alternative for each question, a higher score for the critical alternative indicates that it was selected more often as being the "correct" answer. In other words, *higher* scores represent *higher* or *greater* selection of the critical alternative.

Selection performance of the critical alternatives for the ordered and random component versions of the critical questions was examined in the same three ways described in Experiment 1. In the first set of analyses, using subject means as the random variable (averaged across questions), subjects selected the critical alternatives from the ordered versions (M = .293) significantly more often than would be expected by chance selection (0.25), t (55) = 2.01, p < .05. Similarly, subjects selected the critical alternative from the random version (M = .323) significantly more often than chance, t (55) = 3.30, p < .01. There was no significant difference between the selection scores of the ordered and random scores, t (55) = 1.21, p > .05.

The second set of analyses averaged across subjects and used the critical alternative means as the random variable. Comparisons of the ordered and random versions with chance failed to reach significance, t (19) = 1.26, p > .05, and t (19) = 1.70, p > .05, respectively. The comparison between the ordered and random versions was also not significant, t (19) = .64, p > .05.

In the third set of analyses each of the 20 critical questions was examined individually using the subjects' raw scores as the random variable. Of the 20 critical items, 10 of the ordered versions were in the direction of being selected more often than expected by chance, and 6 of these were significant (p' s < .05). Similarly, 12 of the random versions were in the direction of being selected more often than expected by chance, and 5 of these were significant (p < .05).

With the exception of analyses using the critical alternative means as the random variable, the results of this experiment provide further support for the response bias effect shown in Experiment 1: that the prototypic/critical alternative is chosen more often than would be expected by chance selection.

GENERAL DISCUSSION

The present research sought to determine whether the construction of multiple-choice alternatives based around a critical target answer would facilitate the selection of the target answer. Subjects were given a 60 question multiple-choice test in which 40 of the items had valid answers to legitimate questions. The other 20 questions, the critical questions, asked nonsense, but appeared legitimate. Each had a prototypic/critical alternative around which the other three alternative answers (the distractors) had been systematically derived. This procedure made one of the alternatives (the critical one) more similar to the other alternatives than they were to each other. Analyses examined whether selection of the prototypic/critical alternative was greater than would be expected by mere chance/random selection.

Two experiments were performed. In the first experiment subjects assigned ranking scores to the question alternatives and in the second experiment subjects selected only a single alternative that best answered each question. The results of both experiments produced evidence of a response bias effect. The results show that there is a greater likelihood of selecting the critical target alternatives than would be expected by chance alone. This effect was shown for both the ordered (spatially-arranged) version and the random (scrambled) version and there was no significant difference between these two versions.

Further, the results of Experiment 1 suggest that low ability subjects do not show the response bias effect when the critical alternatives were embedded in sets of distractors with the components are randomly arranged (scrambled). Thus, the spatial arrangement of the components seems to moderate the bias effect in less adept test-takers. One possible contributor to this is that the similarity of the critical alternative to the other alternatives is not as visually apparent in the random component version as in the ordered component version. As briefly argued in the discussion of Experiment 1, the ability to use information provided in the set of alternatives to aid in selection decisions is apparently a form of test-wiseness. Evidently, low ability subjects are not as sophisticated as high ability subjects in this regard.

The results indicate that the bias effect reported by Wogalter and Jensen (1986) and Laughery, Jensen, and Wogalter (in press) is not strictly confined to visual-spatial stimuli but also holds for verbal multiple-choice test items as well. The finding that the response bias effect also occurs for sets of alternatives where the components are randomly arranged (excepting in low ability subjects) suggests that the cognitive processing involved is not limited to visual-spatial imaging of the set of alternatives to form a prototypic image which is then used to match to the critical alternative. The randomly arranged version of the alternatives would seem to hinder this kind of process, and thus it appears that the effect may, in part, involve semantic processes as well as visual/imaging processes. Low ability subjects may use only a visual-spatial imaging process when making their determination; this is sufficient to select the critical alternatives for the ordered but not for the random component version of the questions.

The present results have implications for multiple-choice test construction procedures. They suggest that test-makers should avoid constructing distractor alternatives around a correct alternative because the information provided in the set of alternatives may influence test-takers to select the target alternative even in cases where they lack the knowledge addressed by the question. One possible way to avoid this problem might be to construct the distractors to resemble not just the target but the other distractor items as well.

REFERENCES

Laughery, K. R., Jensen, D. G., & Wogalter, M. S. (in press). Response bias with prototypic faces. In M. M. Gruneberg, P. E. Morris, & R. N. Sykes (Eds.), *Practical Aspects of Memory*. New York: John C. Wiley.

Metfessel, N. S., & Sax, G. (1958). Systematic biases in the keying of correct responses on certain standardized tests. *Educational and Psychological Measurement, 18*, 787-790.

Strang, H. R. (1977). The effects of technical and unfamiliar options on guessing on multiple-choice test items. *Journal of Educational Measurement, 14*, 253-259.

Wilcox, R. R. (1981). Analyzing the distractors of multiple-choice test items or partitioning multinomial cell probabilities with respect to a standard. *Educational and Psychological Measurement, 41*, 1051-1068.

Wogalter, M. S., & Jensen, D. G. (1986). Most similar is different: Response bias in lineups. *Proceedings of the Human Factors Society 30th Annual Meeting* (pp. 725-728). Santa Monica, CA: The Human Factors Society.

Wood, D. A. (1960). *Test Construction* . Columbus, Ohio: Charles E. Merrill Books.